CERN openlab

POSEIDON

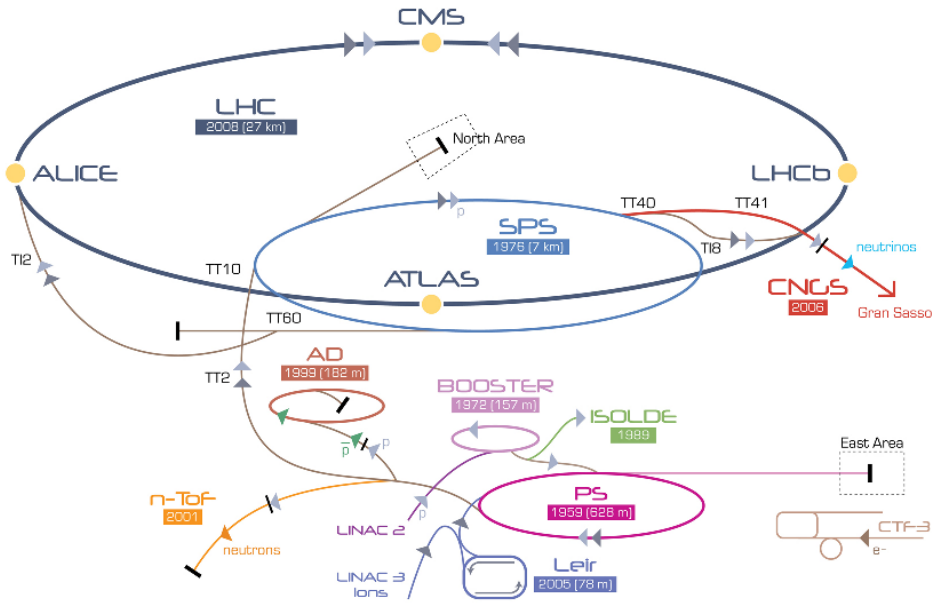
Analyzing the secrets of the Trident Node Monitoring Tool

Supervisors: David Smith, Servesh Muralidharan

Leonardo Kuffo – Escuela Superior Politecnica del Litoral (ESPOL, Ecuador)

IT – DI – WLCG: Understanding Performance

15/08/2018


CMS
LHC
2008 (27 km)
North Area
ALICE
LHCb
TT40
TT41
SPS
1976 (7 km)
TI8
neutrinos
TI2
TT10
ATLAS
CNGS
2006
Gran Sasso
TT60
AD
1999 (182 m)
TT2
BOOSTER
1972 (157 m)
ISOLDE
1989
East Area
n-ToF
2001
PS
1959 (628 m)
CTF-3
neutrons
LINAC 2
LINAC 3
Ions
Leir
2005 (78 m)

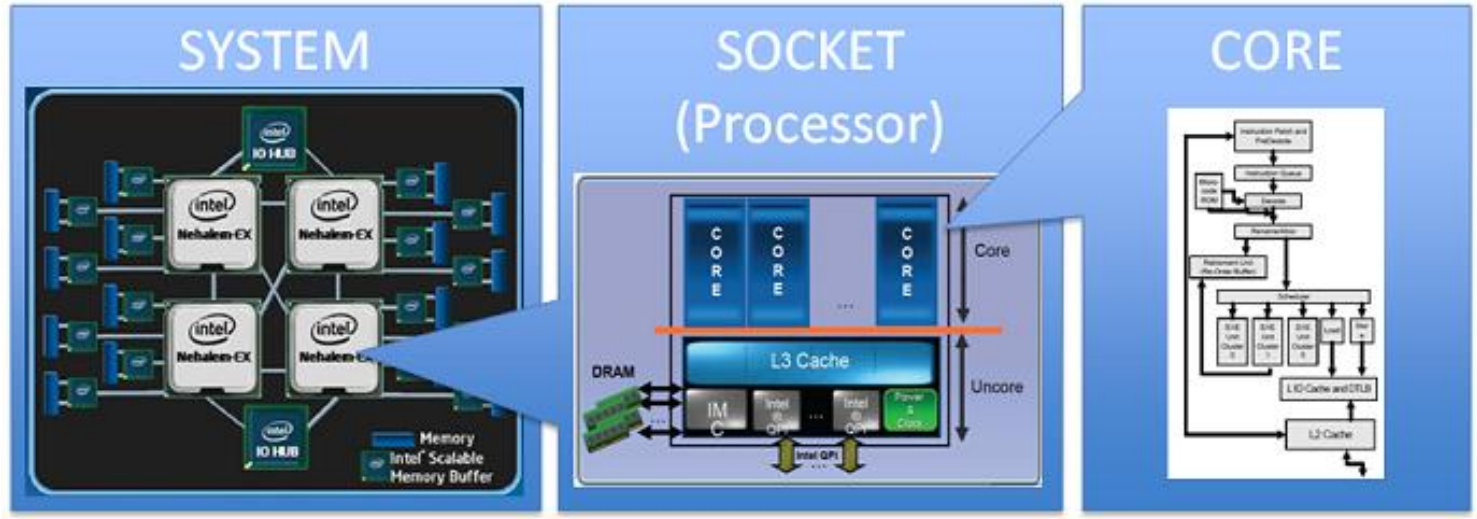

DATA
PACKAGE

Does it do the correct things?

How good is it?

Efficiency / Good Style

Memory Usage / CPU utilization



Intel® Performance Counter Monitor - A better way to measure CPU utilization
By Thomas Willhalm (Intel), Roman Dementiev (Intel), Patrick Fay (Intel)

Trident

Monitoring and Load characterization tool

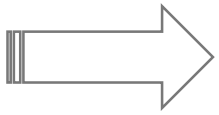
Trident monitors the relevant ***hardware and software*** counters throughout the excecution of an application at the node level, such that it does not induce significant overhead.

Not limited only to monitor CPU Usage or Memory utilization

Currently collecting **metrics** sush as: Memory bandwidth, core utilization, active processor cycles, etc.

Is my application performing well?

+ TRIDENT



Counters Data

POSEIDON

(I hope you get why Poseidon… because I just got it yesterday)

Poseidon - Leonardo Kuffo

POSEIDON

Transforming Trident Data into Knowledge

← Guy who must improve applications performance…

Objective

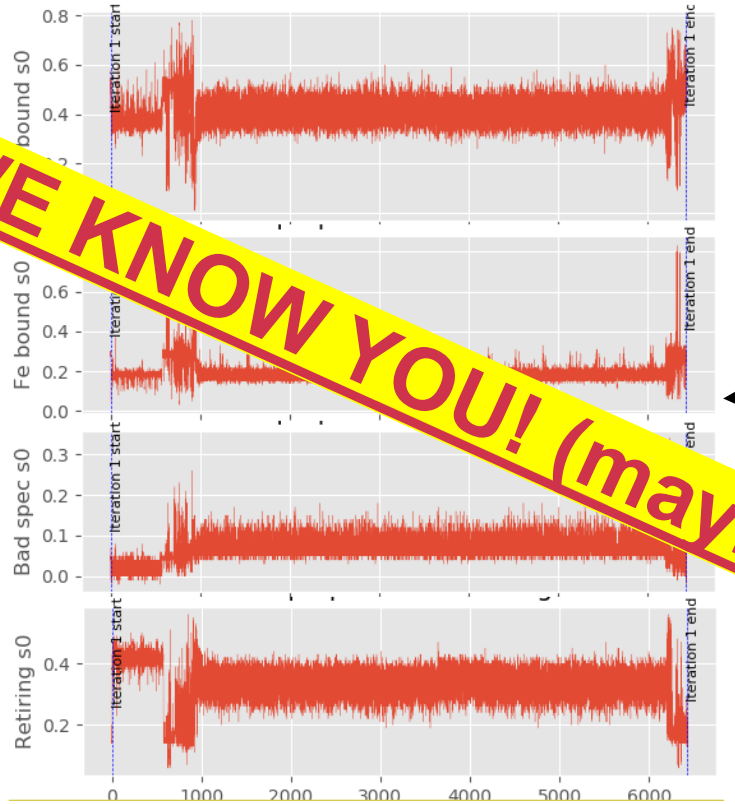
Provide understandable **feedback** to physicist and programmers in how their applications **performs** under different workloads.

Importance

If we understand what is happening, we can improve.

HOW POSEIDON WORKS

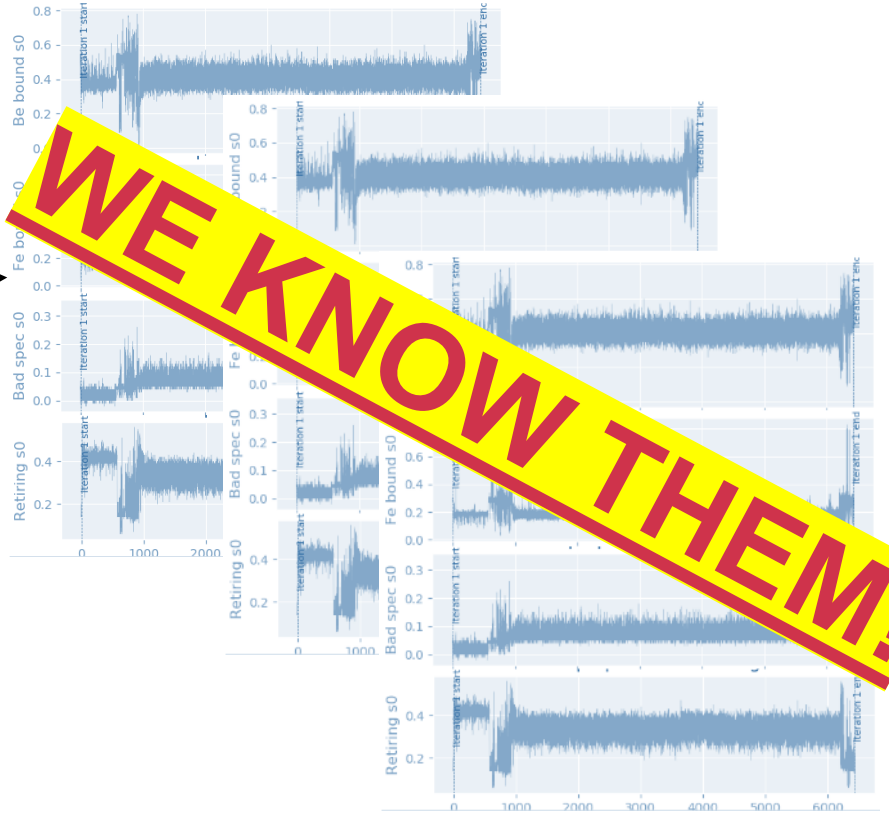
Trident Metrics of an application



WE KNOW YOU! (maybe)

Match

Trident Metrics of Benchmarks



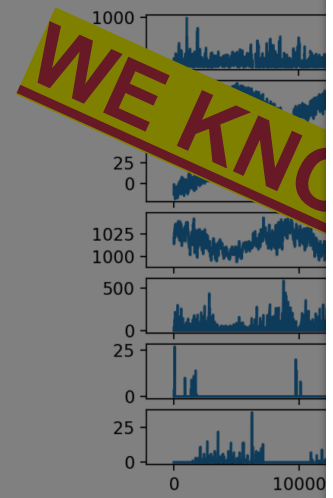
WE KNOW THEM!

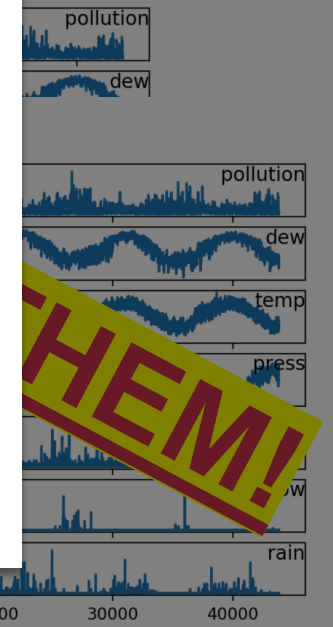
HOW POSEIDON WORKS

We can attack our problem as a **Multi-variate Time series classification** problem.

The benchmarks are our **classes**. At the end, the application workflow will be **matched** to benchmarks.

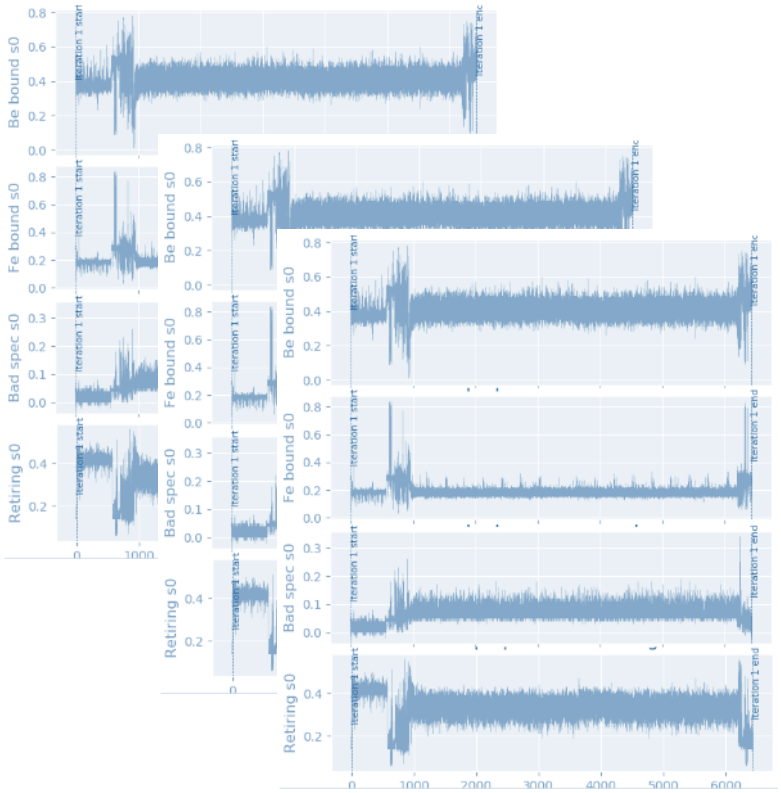
benchmark.





POSEIDON PIPELINE

Trident Metrics of Benchmarks



Train a classification model

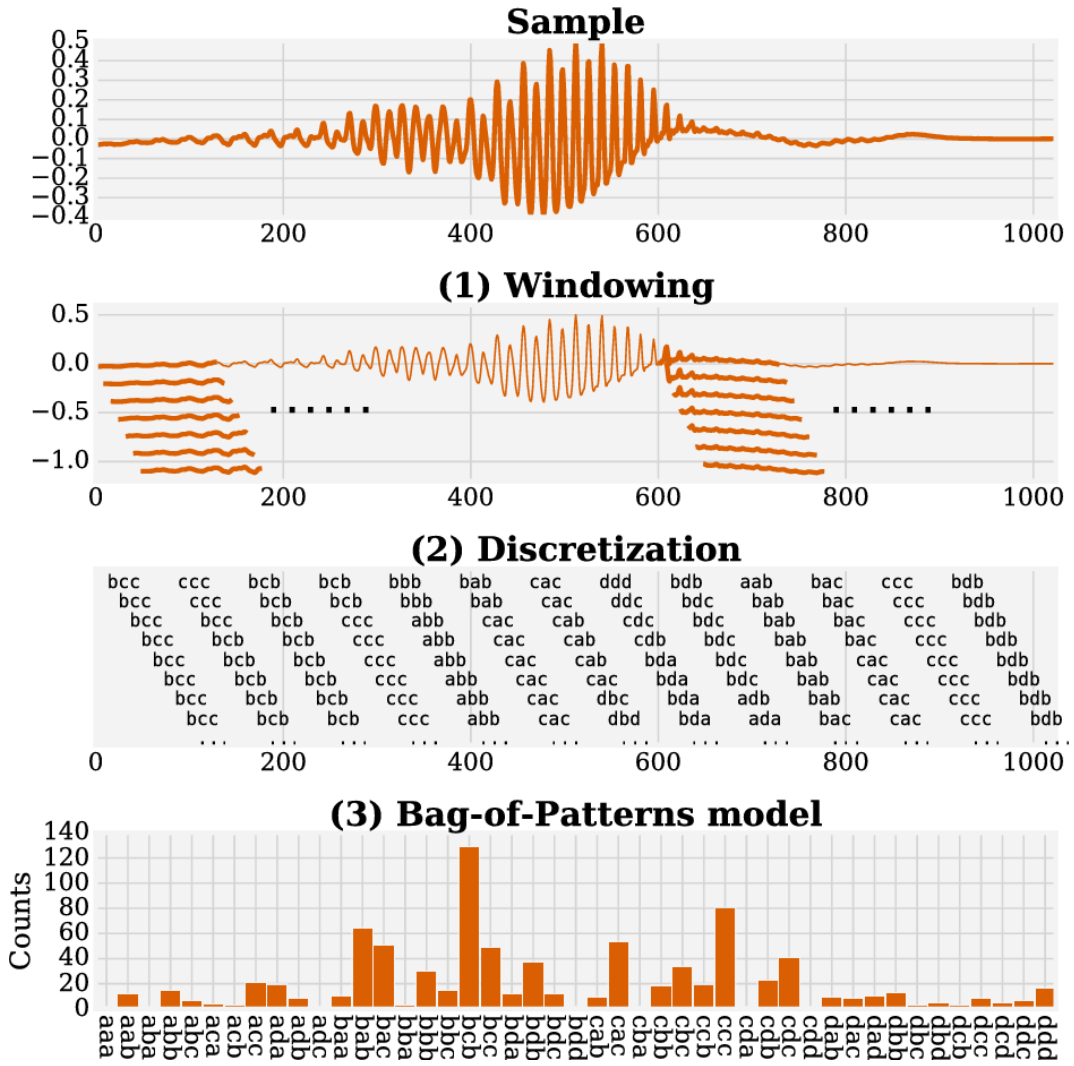
WEASEL TRANSFORMATION

&

Support Vector Machine (SVM)

WEASEL TRANSFORMATION

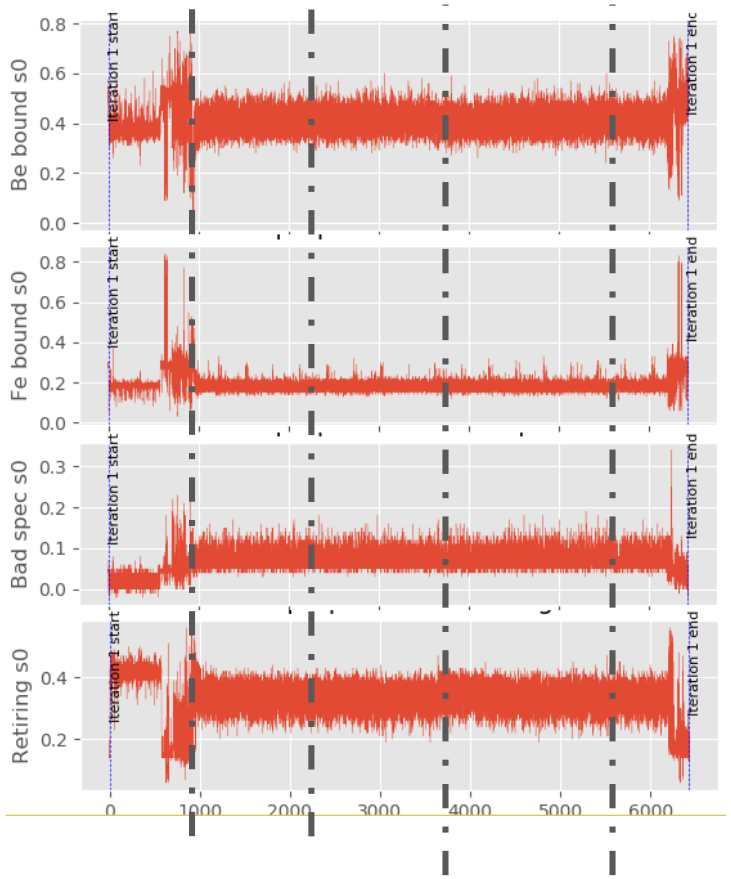
Word ExtrAction for time SEries cLassification



Schäfer, P., & Leser, U. (2017, November). Fast and accurate time series classification with weasel. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 637-646). ACM.

POSEIDON PIPELINE

Trident Metrics of an application



Let the model find the degree of **similarity** with each benchmark in the repository.

of any segment

BINARY SEGMENTATION

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, *42*(6), 2243-2281.

RESULTS & FUTURE WORK

96.36% of accuracy on evaluation data

Test and evaluate POSEIDON results on actual application workflows.

Poseidon can turn into a handy software tool for people who write code at CERN to understand and improve their code at a very low level in a quick fashion.

Special Thanks to: Servesh, David, Markus, Carmen, Family, Dog & Friends.

Thanks Everyone

lkuffo@espol.edu.ec

Leonardo Kuffo

lkuffo

Project Progression Available Online:
https://docs.google.com/document/d/1VgfBxB82IjWcZ9I6kn0jBdWfP82tjIpb5ccVQclrIvE/edit?usp=sharing

References

- Schäfer, P., & Leser, U. (2017). Multivariate time series classification with weasel+ muse. arXiv preprint arXiv:1711.11343.

- Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. Knowledge and information systems, 51(2), 339-367.

- Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association, 107(500), 1590-1598.

- Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. The Annals of Statistics, 42(6), 2243-2281.

- Lin, J., Khade, R., & Li, Y. (2012). Rotation-invariant similarity in time series using bag-of-patterns representation. Journal of Intelligent Information Systems, 39(2), 287-315.

- Schäfer, P., & Leser, U. (2017, November). Fast and accurate time series classification with weasel. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 637-646). ACM.